



Systematic analysis identifies three-lncRNA signature as a potentially prognostic biomarker for lung squamous cell carcinoma using bioinformatics strategy

Jing Hu^{1,2#}, Lutong Xu^{3#}, Tao Shou^{1,2}, Qiang Chen⁴

¹Department of Medical Oncology, The First People's Hospital of Yunnan Province, Kunming 650032, China; ²Department of Medical Oncology, The Affiliated Hospital of Kunming University of Science and Technology, Kunming 650032, China; ³Faculty of Life Science and Technology, Kunming University of Science and Technology, Kunming 650500, China; ⁴Faculty of Animal Science and Technology, Yunnan Agricultural University, Kunming 650201, China

Contributions: (I) Conception and design: Q Chen, T Shou; (II) Administrative support: Q Chen, T Shou; (III) Provision of study materials or patients: J Hu, L Xu; (IV) Collection and assembly of data: J Hu, L Xu, Q Chen; (V) Data analysis and interpretation: J Hu, L Xu; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

[#]These authors contributed equally to this work.

Correspondence to: Qiang Chen. Faculty of Animal Science and Technology, Yunnan Agricultural University, No. 95 of Jinhei Road, Kunming 650201, China. Email: cq@sjtu.edu.cn; Tao Shou. Department of Medical Oncology, The First People's Hospital of Yunnan Province, No. 157 of Jinbi Road, Kunming 650032, China. Email: taoshou_yn@163.com.

Background: Lung squamous cell carcinoma (LUSC) is the second most common histological subtype of lung cancer (LC), and the prognoses of most LUSC patients are so far still very poor. The present study aimed at integrating lncRNA, miRNA and mRNA expression data to identify lncRNA signature in competitive endogenous RNA (ceRNA) network as a potentially prognostic biomarker for LUSC patients.

Methods: Gene expression data and clinical characteristics of LUSC patients were retrieved from The Cancer Genome Atlas (TCGA) database, and were integratedly analyzed using bioinformatics methods including Differentially Expressed Gene Analysis (DEGA), Weighted Gene Co-expression Network Analysis (WGCNA), Protein and Protein Interaction (PPI) network analysis and ceRNA network construction. Subsequently, univariate and multivariate Cox regression analyses of differentially expressed lncRNAs (DElncRNAs) in ceRNA network were performed to predict the overall survival (OS) in LUSC patients. Receiver operating characteristic (ROC) analysis was used to evaluate the performance of multivariate Cox regression model. Gene expression profiling interactive analysis (GEPIA) was used to validate key genes.

Results: WGCNA showed that turquoise module including 1,694 DElncRNAs, 2,654 DE mRNAs as well as 113 DE miRNAs was identified as the most significant modules ($cor=0.99$, $P<1e-200$), and differentially expressed RNAs in the module were used to subsequently analyze. PPI network analysis identified *FPR2*, *GNG11* and *ADCY4* as critical genes in LUSC, and survival analysis revealed that low mRNA expression of *FPR2* and *GNG11* resulted in a higher OS rate of LUSC patients. A lncRNA-miRNA-mRNA ceRNA network including 121 DElncRNAs, 18 DE miRNAs and 3 DE mRNAs was established, and univariate and multivariate Cox regression analysis of those 121 DElncRNAs showed a group of 3 DElncRNAs (*TTY16*, *POU6F2-AS2* and *CACNA2D3-AS1*) had significantly prognostic value in OS of LUSC patients. ROC analysis showed that the area under the curve (AUC) of the 3-lncRNA signature associated with 3-year survival was 0.629.

Conclusions: The current study provides novel insights into the lncRNA-related regulatory mechanisms underlying LUSC, and identifying 3-lncRNA signature may serve as a potentially prognostic biomarker in predicting the OS of LUSC patients.

Keywords: Lung squamous cell carcinoma (LUSC); weighted gene co-expression network analysis (WGCNA); competing endogenous RNA network; long non-coding RNA; overall survival (OS)

Submitted Apr 10, 2019. Accepted for publication Aug 29, 2019.

doi: 10.21037/tlcr.2019.09.13

View this article at: <http://dx.doi.org/10.21037/tlcr.2019.09.13>

Introduction

Lung cancer (LC) with more than 2 million new cases and 1.7 million deaths in 2018, is the most commonly diagnosed malignancies and the top leading cause of cancer death worldwide (1). Non-small cell lung cancer (NSCLC) is the most common pathological type, and accounts for approximately 85% of all LCs (2). Among NSCLCs, lung squamous cell carcinoma (LUSC) with more than 400,000 new cases per year is the second most common type of NSCLCs, and accounts for 20–30% of NSCLCs (3,4). Due to the lack of sensibility to radiotherapy, surgery and chemotherapy are the main methods to combat LUSC (5). Despite advances in treatment methods of LUSC, the 5-year overall survival (OS) rate of LUSC patients with clinical I and II stages is about 40%, and that of LUSC patients with clinical III and IV stages less than poor 5% (6). A diagnosis for LUSC at an early stage contributes to offering a favorable prognosis, and the 5-year OS rate will significantly increase to 70–90% (7). Currently, disease stage and histological grade are still the basic methods for evaluating the diagnosis and prognosis of LUSC. However, it is limited to evaluate the predictive value in diagnosing early-stage LUSC by clinical and pathological symptoms, and it is difficult to predict the clinical outcomes due to the heterogeneity of LUSC. In addition, the molecular mechanisms underlying LUSC still remain unclear. Therefore, it is vital to further explore the molecular mechanisms, and identify potentially molecular diagnostic and prognostic markers and/or therapeutic targets to combat LUSC.

Smoking is the major risk factor for LUSC, which causes high number of mutations such as *TP53*, *CDKN2A* and *PIK3CA* (8). Although molecularly targeted therapies directly against those driver genes are an effective method to combat LUSC, development of molecularly targeted therapies has been less rapid for LUSC with respect to lung adenocarcinoma (LUAD), and molecularly targeted therapies are still not standard among LUSCs. Even though some genes such as *DDR2*, *AKT* and *FGFR* were considered as potentially molecular targets, few effective targeted therapies for LUSC are so far approved, and targeted therapies developed against LUAD are largely ineffective

against LUSC (9,10). Recent efforts have focused on discovering novel, clinically actionable cancer driver genes by multi-omics analysis (10,11). However, due to the accumulation of large number of passenger mutations from prolonged exposure to carcinogens, identifying novel driver genes is challenging (8).

Long non-coding RNAs (lncRNAs) are an important class of non-coding RNAs (ncRNAs) with over 200 nucleotides long (12), and are widely reported to play roles as important regulatory components in tumor initiation and progression (13). However, little knowledge about the roles of lncRNAs is known in tumor biology. Nevertheless, due to the high tissue specificity of lncRNA expression, lncRNAs have been considered as potentially valuable diagnostic and prognostic biomarkers and/or therapeutic targets of some malignant tumors (14–17). Competitive endogenous RNA (ceRNA) hypothesis is so far considered as the foundation of playing the regulatory role for lncRNAs (18). lncRNA as ceRNA and mRNA can competitively bind microRNA (miRNA) by miRNA-response-elements (MREs) to play key roles in various biological processes such as tumor formation and metastasis (19). For instance, Fan's team constructed lncRNA-miRNA-mRNA ceRNA network based on ceRNA theory to identify that a 4-lncRNA signature played roles in breast cancer (20). Liu and the colleagues revealed functional lncRNAs in gastric cancer by integrative analysis of dysregulated lncRNA-related ceRNA network (21). In addition, systematic analysis of lncRNA-related ceRNA network have been reported in many other cancers including thyroid carcinoma (22), glioblastoma (23), prostate cancer (24), and ovarian cancer (25). Collectively, the lncRNA-related ceRNA network findings revealed the lncRNA-associated regulatory mechanisms by miRNA-mediated lncRNA and mRNA interaction in cancer initiation and progression. For lncRNA-related ceRNA network of LUSC, several studies have reported some lncRNAs associated with LUSC development and clinical outcome (9,26,27). However, the findings reported from different studies are inconsistent due to differentially limited samples. The Cancer Genome Atlas (TCGA, 2018, the seventh edition of TNM staging criteria, <https://cancergenome.nih.gov/>) is a publically available cancer-related database which provides large-

scale multi-dimensional molecular profile data related to over 30 cancer types. Using LUSC-related data with large sample size from TCGA database to identify LUSC-specific lncRNAs can increase statistical reliability of the current study. Another key issue is that most reported studies didn't consider the co-expression characteristic of genes in ceRNA network, which may be difficult to target the key lncRNAs. Weighted gene co-expression network analysis (WGCNA) is a systematic biology method to describe the correlation patterns among genes across samples, and using WGCNA can find modules of highly correlated genes and identify key genes (28). Presently, WGCNA has been successfully applied in various biological contexts including cancer such as LC (29), bladder cancer (30), and colon cancer (31). Systematic analysis of lncRNA-miRNA-mRNA ceRNA network based on WGCNA contributes to insight into novel molecular mechanisms involved in LUSC. While, very little information is available on LUSC ceRNA based on WGCNA.

In this work, LUSC-related RNA-seq data were retrieved from the TCGA database, and were systematically integrated and analyzed based on bioinformatics methods including Differentially Expressed Gene Analysis (DEGA), WGCNA, Protein and Protein Interaction (PPI) network, Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses, ceRNA network, and centrality analysis. Subsequently, survival analysis was performed to identify the key lncRNA signature in lncRNA-related ceRNA network as potentially valuable prognosticator associated with OS of LUSC patients. This study provides novel insights into molecular mechanisms underlying LUSC, and identifying novel lncRNAs may serve as potentially diagnostic and prognostic biomarkers and/or therapeutic targets to combat LUSC.

Methods

The flow chart of systematic bioinformatics analysis in the current study was showed in *Figure 1*.

RNA-seq data and patient information collection

LUSC-related RNA-seq data were retrieved from TCGA database portal (2018, <https://cancergenome.nih.gov/>), and the inclusion criteria of RNA-seq data were included as follows: (I) histological diagnosis for LUSC; (II) except LUSC no other malignancy; (III) data with complete clinical information. A total of 475 LUSC tissues and

38 non-LUSC normal lung tissues RNA-seq data were retrieved. Among, 38 non-LUSC normal lung tissues were from the normal lung tissues nearby LUSC tissues of 38 LUSC patients. The clinical information of all LUSC samples patients was simultaneously obtained from TCGA database. Sample clustering was used to detect outlier, and 2 LUSC tissues samples were excluded. Finally, a total 473 LUSC-related and 38 non-LUSC normal lung tissues RNA-seq data was included. All RNA-seq data were approved by the Institutional Review Board of the relevant participating institutions and by the National Cancer Institute of NIH. No approval from the ethics committee was required. The present study meets the requirements of data usage and publishing from TCGA database. The clinical characteristics for LUSC patients were listed in *Table 1*.

RNA-seq data preprocessing and DEGA

All raw RNA-seq data were subjected to normalization using the trimmed mean of M-values method.

EdgeR package in Bioconductor project (version 3.8, <http://www.bioconductor.org/>) was used to screen the differentially expressed lncRNAs (DElncRNAs), differentially expressed miRNAs (DEmiRNAs) and differentially expressed mRNAs (DEmRNAs) between LUSC tissues and non-LUSC lung tissues samples (32). $|\log(\text{fold change})|$ ($|\logFC|$) > 2 and statistical P value < 0.01 ($P < 0.01$) were set as cut-off criteria. The volcano plots of RNA expression were visualized using the ggplot2 package (version 3.1.0, <https://github.com/tidyverse/ggplot2>). The heat map of differentially expressed RNAs (DERNAs) was plotted using the pheatmap package (version 1.0.10, <https://CRAN.R-project.org/package=pheatmap>).

WGCNA

WGCNA was used to construct the gene co-expression network among DElncRNAs, DEmiRNAs and DEmRNAs, and identify co-expression gene modules. WGCNA package (version 1.13) based on R was used to perform WGCNA (28). The expression matrix was firstly converted into an adjacency matrix, and an unsupervised co-expression relationship was constructed based on the adjacency matrix using Pearson correlation coefficients for gene pairs. The correlation adjacency matrix was strengthened by power $\beta = 10$ (soft threshold), and the power parameter was selected based on scale-free topology criterion. Secondly, the

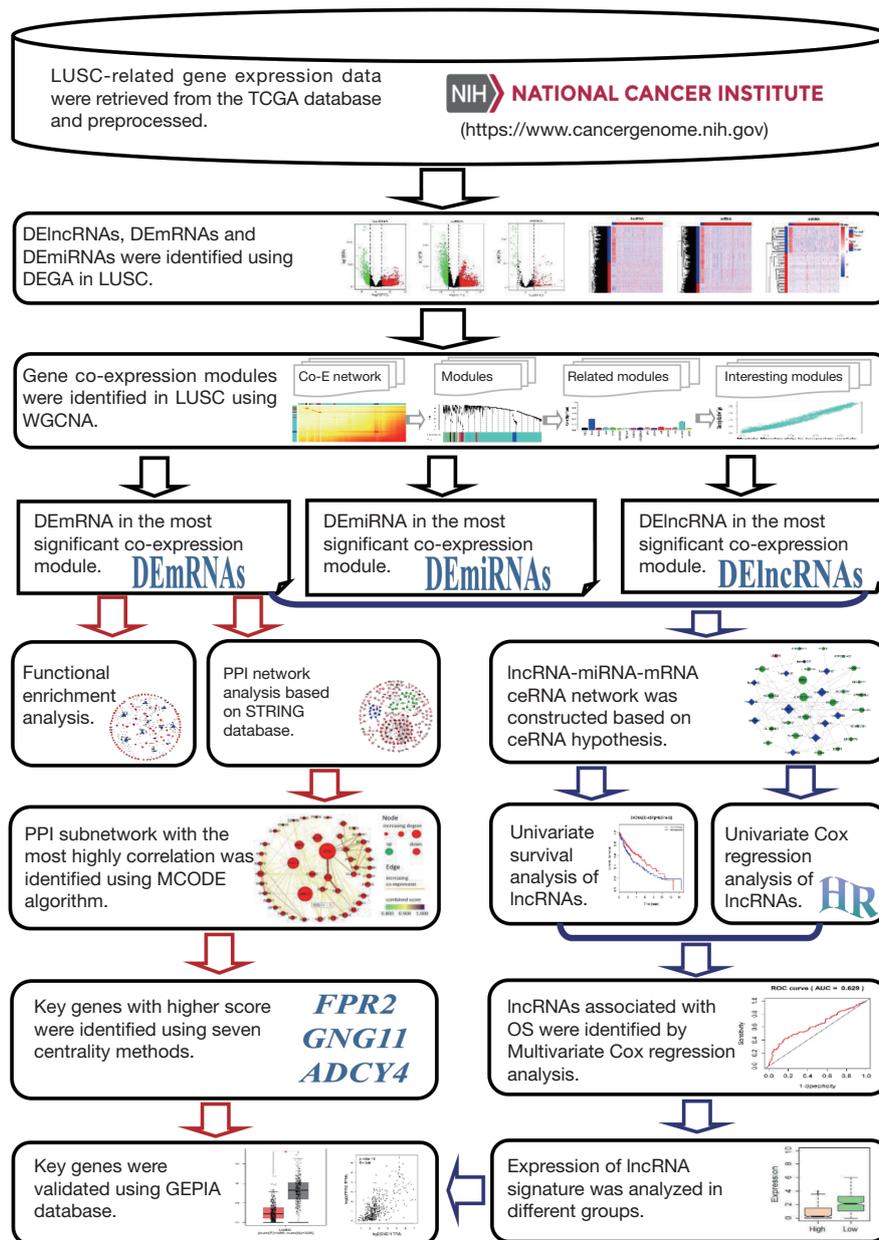


Figure 1 Flow chart of bioinformatics analysis. LUSC, lung squamous cell carcinoma; DEGA, differentially expressed gene analysis; DElncRNA, differentially expressed lncRNA; WGCNA, weighted gene co-expression network analysis; PPI, protein and protein interaction; ceRNA, competitive endogenous RNA; MCODE, molecular complex detection; GEPIA, gene expression profiling interactive analysis.

adjacency matrix was converted into a topological matrix, and topological overlap measure (TOM) was used to measure the correlation of gene pairs. According to TOM-based dissimilarity, average linkage hierarchical clustering was performed to classify genes with coherent expression

profiles into gene modules. Dynamic cutting algorithm was used to identify gene modules from the system cluster tree, and the modules with 95% similarity were merged into a module. Module eigengene (ME) was defined as the first principal components, and was the representative in

Table 1 The predictive values of clinical features and risk score

Variable	Patients (N)	Univariate analysis		Multivariate analysis	
		HR (95% CI)	P	HR (95% CI)	P
Gender (male/female)	371/130	0.955 (0.694–1.316)	0.780	1.095 (0.793–1.512)	0.580
Pathologic stage (I–II/III–IV)	406/95	1.226 (0.894–1.681)	0.206	0.805 (0.586–1.105)	0.179
Smoking (yes/no)	222/279	0.778 (0.593–1.020)	0.069	1.301 (0.991–1.710)	0.058

HR, hazard ratio; CI, confidence interval.

module genes. Module membership (MM) was defined as the correlation between ME and gene module. Gene significance (GS) was indexed by log₁₀ transformation of the p value of *t*-test measuring differential expression between LUSC and non-LUSC tissues. Module significance (MS) was defined as the average GS for all the genes in the module. More detailed description of WGCNA is presented by original article (28).

The genes in the most significant module were selected to further analyze.

PPI network construction and analysis

The interactive relationships among differentially expressed genes (DEGs) encoding proteins in the most significant gene co-expression module was analyzed by constructing a PPI network. The interactive information among DEGs encoding proteins was retrieved from online STRING database (version 11.0, <https://string-db.org/>) (33). The gene pairs with a combined score \geq the highest confidence 0.9 were used to construct the PPI network. Cytoscape software (version 3.7.0, <http://www.cytoscape.org/>) was used to construct and visualize the interactive relationships among genes in whole PPI network (34). The subnetwork (highly correlated module) was extracted from whole PPI network using Molecular COMplex DETection (MCODE) algorithm based on topological properties of whole PPI network, and a plugin MCODE (version 1.5.1) in Cytoscape was used to perform MCODE analysis (35). The threshold parameters were more severely set for Degree Cutoff=4, Node Score Cutoff=0.6, K-Core=4 and Max. Depth=100. The most significantly correlated module was used to next analyze.

Gene function analysis

GO and KEGG pathway enrichment analyses of DE mRNAs in PPI network were implemented using the

clusterProfiler package (version 3.10.1), and an adjusted $P < 0.05$ was considered to significantly enrich (36). GO analysis included Biological Process (BP), Cellular Component (CC) and Molecular Function (MF).

Identification of key genes and validation

Key genes in PPI network were identified using Centrality analyses. Centrality analyses including Subgraph Centrality, Degree Centrality, Eigenvector Centrality, Betweenness Centrality, Network Centrality, Information Centrality and Closeness Centrality were performed using a plugin CytoNCA (version 2.1.6) in Cytoscape (37). The genes with higher Centrality scores were identified as key gene, and essential genes were identified as intersecting genes of key genes obtained by seven Centrality methods. Gene Expression Profiling and Interactive Analyses (GEPIA, <http://gepia.cancer-pku.cn>) database is an interactive web server for analyzing gene expression data of tumors and normal tissues from TCGA and genotype-tissue expression database (38,39), and was used to validate the expression of essential genes in PPI network between LUSC and non-LUSC lung normal tissues and analyze the association with OS in LUSC patients. In GEPIA database, the expression analysis between LUSC and non-LUSC lung normal tissues was performed using $\log_2(\text{TPM}+1)$, and survival analysis was performed using Log-Rank (LR) test (Mantel-Cox test) for hypothesis test. The LR P value, cox proportional hazard ratio (HR) and the 95% confidence interval (CI) information were calculated.

Construction of lncRNA-miRNA-mRNA ceRNA network

Co-expression genes in the most significant module were used to construct lncRNA-miRNA-mRNA ceRNA network, and ceRNA network was established based on ceRNA hypothesis that lncRNA directly bind miRNA by acting as miRNA sponge to indirectly regulate the

function of mRNA (18). A ceRNA network was constructed according to the following steps: (I) three types of DERNA in the most significant module were kept; (II) the potential target miRNAs of DElncRNAs, as well as lncRNAs and miRNAs interaction relationships were predicted using the online miRcode tool (miRcode 11, <http://www.mircode.org/>); (III) the potential target mRNAs of DEmiRNAs were predicted using the online tools including TargetScan (release 7.2, <http://www.targetscan.org/>), miRDB (version 5.0, <http://mirdb.org/>) and miRTarBase (release 7.0, <http://mirtarbase.mbc.nctu.edu.tw/>); (IV) based on ceRNA hypothesis, intersecting miRNAs negatively regulated by lncRNAs and mRNAs were selected to construct ceRNA network. A ceRNA network was constructed and visualized using an open-source Cytoscape software (34).

LUSC-specific prognostic lncRNA signatures identification

The associations between DElncRNAs in ceRNA network and OS in LUSC patients were evaluated using Kaplan-Meier (KM) estimate and Log-rank (LR) test in survival package (version 2.43-3, <https://CRAN.R-project.org/package=survival>) based on R. Group cut off of 50% was set, and LR P value, hazard ratio (HR) and 95% CI were computed. The statistical $P < 0.05$ was considered as the significant association between DElncRNA and OS of LUSC patients. Further, univariate Cox proportional hazards regression model on the basis of survival package in R project was performed to evaluate the association between DElncRNAs in ceRNA network and OS of LUSC patients. The same characteristic parameters as LR method were computed and the same significant P value criterion was set. Overlapping DElncRNAs significantly associated with OS of LUSC patients in LR method and univariate Cox regression analysis was performed multivariate Cox hazards regression model with Stepwise method, and which was implemented using survival package in R project, and multivariate Cox hazards regression model was used to assess the prognostic value for LUSC. DElncRNA combination in the optimal Cox hazard regression model was used to further analyses. The hazards model was established as follows:

$$\text{Risk score} = \text{Exp}_{\text{lncRNA1}} * \text{Coe}_{\text{lncRNA1}} + \text{Exp}_{\text{lncRNA2}} * \text{Coe}_{\text{lncRNA2}} \\ + \text{Exp}_{\text{lncRNA3}} * \text{Coe}_{\text{lncRNA3}} + \dots + \text{Exp}_{\text{lncRNA}_n} * \text{Coe}_{\text{lncRNA}_n}$$

where “Exp” denoted the expression level of lncRNA, and “Coe” represented the regression coefficient from the multivariate Cox regression model (20). According

to the median of above risk scores, LUSC patients were divided into high- and low-risk groups. Receiver operating characteristic (ROC) curve was constructed using survivalROC package (version 1.0.3, <https://CRAN.R-project.org/package=survivalROC>) based on R, and was used to measure the risk prediction rate of lncRNAs between high- and low-risk groups.

Statistical analysis

All statistical analyses were performed using R software. The comparison of gene expression in LUSC tissues and normal tissues was analyzed using Mean \pm Standard Deviation. T-test was used to estimate the statistical difference, and $P < 0.05$ was considered as significant difference between two groups in expression level.

Results

Identification of DElncRNA, DEmRNA and DEmiRNA

After all raw data were normalized, 8,809 lncRNAs, 18,275 mRNAs, and 821 miRNAs were extracted. Using DEGA based on EdgeR package, we identified DElncRNAs, DEmRNAs and DEmiRNAs between LUSC tissues samples and non-LUSC lung tissues samples. According to the $|\log_{2}FC| > 2$ and $P < 0.01$ cut-off criteria, a total of 2,114 DElncRNAs (1,667 up- and 447 down-regulated, <http://fp.amegroups.cn/cms/32d9bc8d31415aaab9827d0601d1fbbd/tlcr.2019.09.13-1.pdf>), 3,485 DEmRNAs (2,324 up- and 1,161 down-regulated, <http://fp.amegroups.cn/cms/a52c08dd78c153449baca4f523871775/tlcr.2019.09.13-2.pdf>), and 175 DEmiRNAs (148 up- and 27 down-regulated, <http://fp.amegroups.cn/cms/a63f404926063d02b2c7c606a1972a92/tlcr.2019.09.13-3.pdf>) were identified. The distributions of DElncRNAs, DEmRNAs and DEmiRNAs were displayed using volcano plots in *Figure 2A*. The expression levels of DElncRNAs, DEmRNAs and DEmiRNAs between LUSC tissues and non-LUSC normal lung tissues were displayed using heat map in *Figure 2B*.

Co-expression module identification

All DERNA including 2,114 DElncRNAs, 3,485 DEmRNAs and 175 DEmiRNAs were used to reconstruct the transcriptional network using WGCNA algorithm. Pearson correlation matrix among genes were converted into a strengthened adjacency matrix by power $\beta = 10$ based

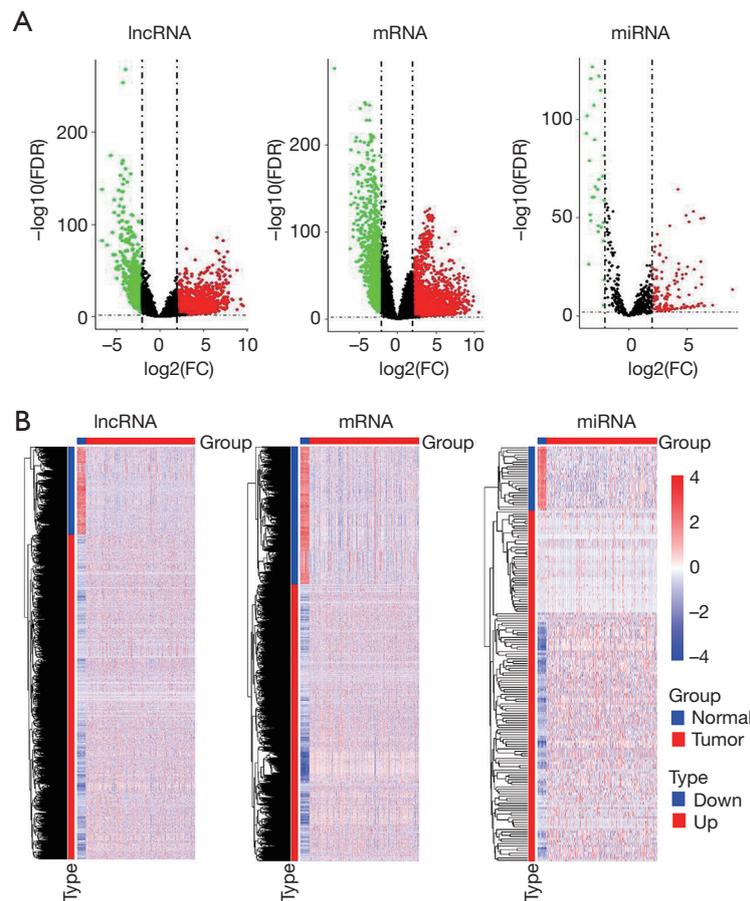


Figure 2 Differentially expressed RNAs in LUSC. (A) Volcano plot of expressions of three types of RNAs. X axis represented the mean expression differences of lncRNAs, mRNAs and miRNAs between LUSC and normal lung tissues samples, and Y axis represented log transformed false discovery rate; (B) heatmap of expressions of three types of differentially expressed RNAs between LUSC and normal lung tissues samples. X-axis represented differentially expressed lncRNAs, mRNAs and miRNAs, and blue represented normal lung tissues samples and red represented LUSC tissues samples. Y-axis represented the samples, and blue represented down-regulated lncRNAs, mRNAs and miRNAs, and red represented up-regulated lncRNAs, mRNAs and miRNAs. The color scale showed the expression values. LUSC, lung squamous cell carcinoma.

on scale-free topology criterion with $R^2=0.94$ (Figure 3A,B). TOM of each gene pair was calculated, and 16 gene co-expression modules were identified by average linkage hierarchical clustering according to TOM-based dissimilarity measure (1-TOM) (Figure 3C). The correlation analysis between ME and LUSC showed that turquoise module (included 1,694 DElncRNAs and 2,654 DEMRNAs as well as 113 DEmiRNAs, <http://fp.amegroups.cn/cms/23ad548482adf5ccf686ce250556033f/tlcr.2019.09.13-4.pdf>) was the most significantly representative module ($cor=-0.9$ and $P=5e-190$) (Figure 3D). The MM in turquoise module possessed the most significant correlation in all modules ($cor=0.99$ and

$P<1e-200$) (Figure 3E), and GS analysis across modules showed that turquoise module had the second highest gene significance with the lowest standard deviation (Figure 3F). Thus, the turquoise module was used for next analyses.

GO and KEGG enrichment analyses

Based on MM, 660 MEs with $|MM|>0.6$ in turquoise module were selected to investigate the functions by GO and KEGG enrichment analyses. GO enrichment results showed that 216 BPs, 58 CCs and 33 MFs were significantly enriched. KEGG enrichment analysis showed that 11

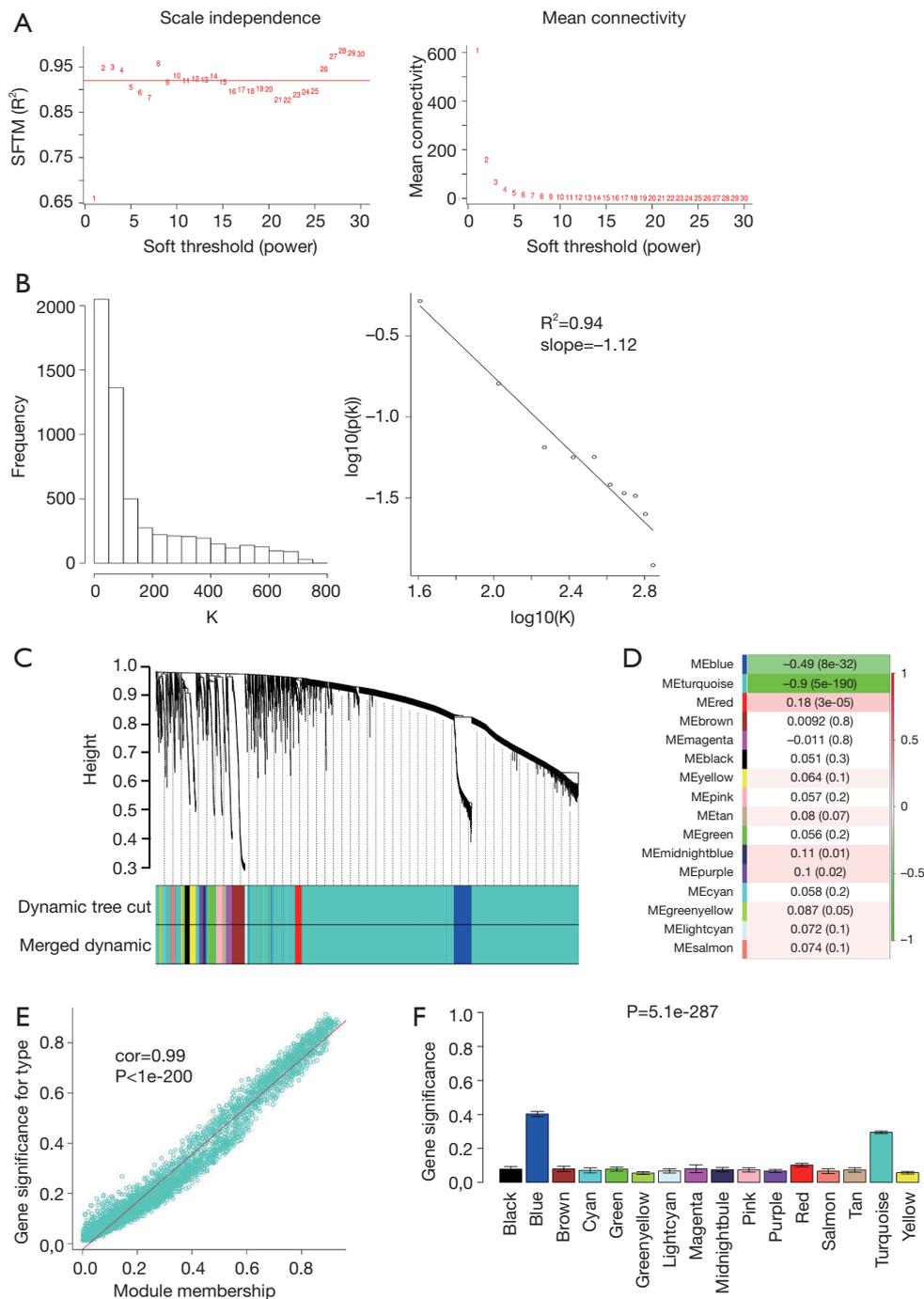


Figure 3 Co-expression network analysis by WGCNA. (A) Analysis of network topology for various soft-threshold powers. (B) The test of property of scale-free network. (C) Identification of LUSC-specific modules. Each short vertical line corresponded to a gene. Each branch represented a expression module of highly interconnected groups of genes. Below the dendrogram, each group of genes was given a color, which indicated its module assignment. Gray suggested that genes were outside all modules. (D) The associations of modules and LUSC. Turquoise module was identified to be the most significant association with LUSC. (E) The associations of turquoise module membership and LUSC. (F) Barplot of mean significance across modules. The higher mean gene significance represented the higher correlation between module and LUSC. LUSC, lung squamous cell carcinoma; WGCNA, weighted gene co-expression network analysis.

Table 2 Top 10 pathways enriched by GO and KEGG analyses

ID	Description	Adj. P value	Count
Biological process (BP)			
GO:0023052	Signaling	8.15E-12	180
GO:0007155	Cell adhesion	4.84E-10	64
GO:0022610	Biological adhesion	4.84E-10	64
GO:0007165	Signal transduction	1.75E-08	116
GO:0007166	Cell surface receptor linked signaling pathway	3.15E-08	88
GO:0042221	Response to chemical stimulus	3.48E-08	96
GO:0006952	Defense response	4.57E-08	54
GO:0006954	Inflammatory response	4.57E-08	36
GO:0023033	Signaling pathway	5.23E-08	123
GO:0023046	Signaling process	5.94E-08	125
Cellular component (CC)			
GO:0044459	Plasma membrane part	3.78E-24	159
GO:0005887	Integral to plasma membrane	2.36E-22	112
GO:0031226	Intrinsic to plasma membrane	2.36E-22	113
GO:0005886	Plasma membrane	1.03E-21	229
GO:0016020	Membrane	6.95E-16	342
GO:0044421	Extracellular region part	1.00E-12	82
GO:0044425	Membrane part	3.79E-12	290
GO:0005615	Extracellular space	8.09E-12	67
GO:0031224	Intrinsic to membrane	6.48E-10	255
GO:0005576	Extracellular region	8.47E-10	123
Molecular function (MF)			
GO:0030246	Carbohydrate binding	2.46E-08	40
GO:0060089	Molecular transducer activity	4.55E-07	120
GO:0004871	Signal transducer activity	4.55E-07	120
GO:0004872	Receptor activity	8.65E-07	100
GO:0005529	Sugar binding	2.39E-05	23
GO:0030247	Polysaccharide binding	2.50E-04	19
GO:0001871	Pattern binding	2.50E-04	19
GO:0005539	Glycosaminoglycan binding	9.16E-04	17
GO:0019199	Transmembrane receptor protein kinase activity	1.17E-03	12
GO:0004888	Transmembrane receptor activity	1.17E-03	70

Table 2 (continued)

Table 2 (continued)

ID	Description	Adj. P value	Count
KEGG pathway			
hsa04514	Cell adhesion molecules (CAMs)	0.0008	18
hsa04610	Complement and coagulation cascades	0.00084	13
hsa04512	ECM-receptor interaction	0.0036	12
hsa05414	Dilated cardiomyopathy (DCM)	0.0055	12
hsa04151	PI3K-Akt signaling pathway	0.009	26
hsa05410	Hypertrophic cardiomyopathy (HCM)	0.009	11
hsa04080	Neuroactive ligand-receptor interaction	0.02	21
hsa04510	Focal adhesion	0.02	17
hsa03320	PPAR signaling pathway	0.0326	9
hsa05144	Malaria	0.0418	7

GO, gene ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes.

KEGG pathways were significantly enriched. Top 10 GO terms and KEGG pathways were showed in *Table 2*, and the most significant BP, CC, MF and KEGG pathway were Signaling (GO:0023052, $P=8.15E-12$), Plasma membrane part (GO:0044459, $P=3.78E-24$), Carbohydrate binding (GO:0030246, $P=2.46E-08$) and Cell adhesion molecules (hsa04514, $P=0.0008$), separately. In order to elucidate the potentially biological complexities, pathway-gene network was constructed to show the relationships between pathways and genes (*Figure 4A*). PI3K-Akt signaling pathway contained the most significant genes in the network, which involved in many cellular functions such as transcription, translation, proliferation, growth, and survival.

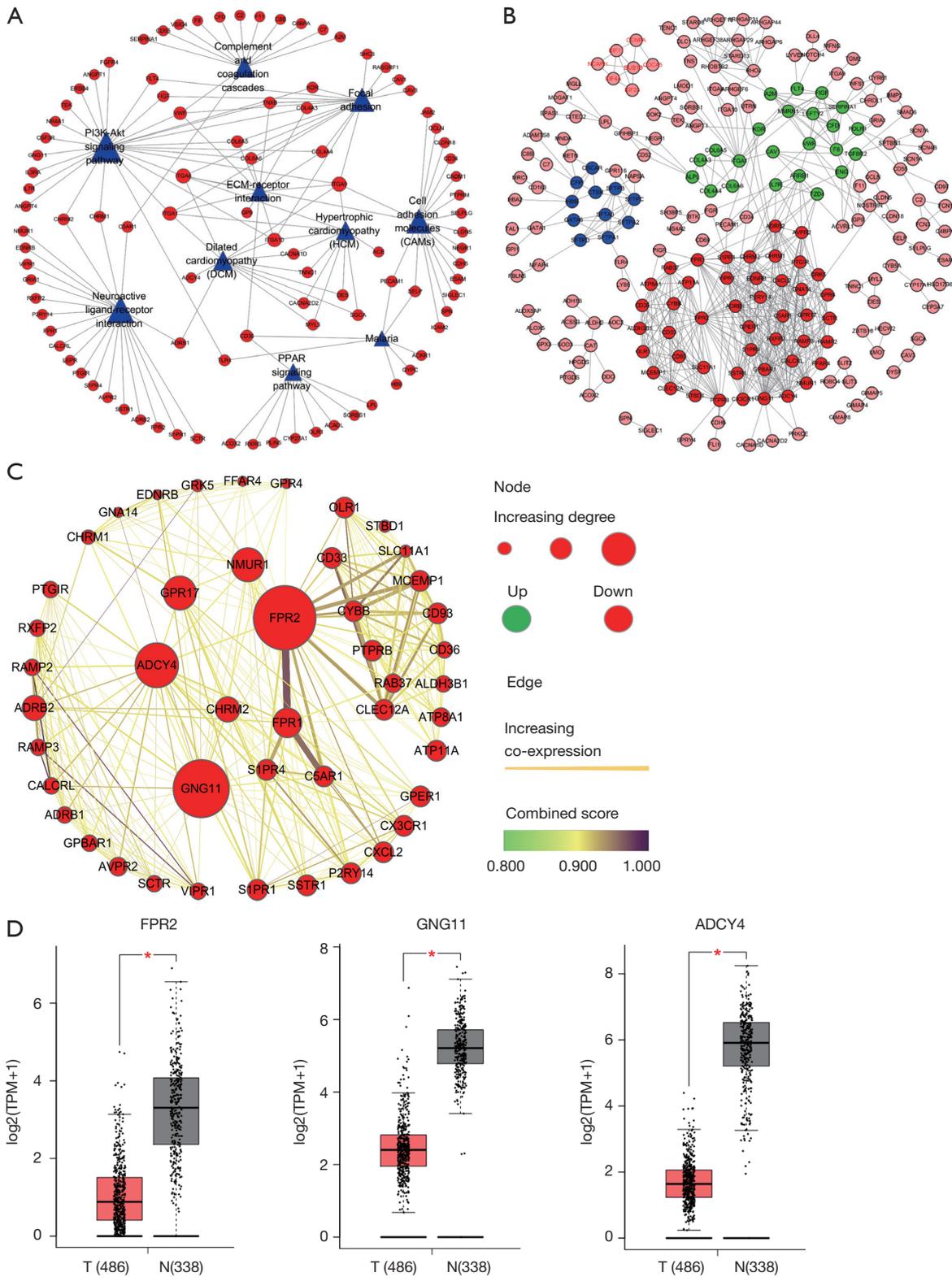
PPI network construction and key gene validation

DEmRNAs (MEs) in the most significant turquoise module were used to perform PPI network analysis. Based on MM, 660 MEs with $|MM|>0.6$ (653 positive correlation and 7 negative correlation) were selected to construct PPI network. At minimum required interaction score = the highest confidence 0.9, a total of 239 among 660 MEs was filtered into PPI network, and a PPI network with 239 nodes and 625 edges was established (*Figure 4B*). Highly correlated module analysis showed that 3 significantly correlated modules were identified in whole PPI network, and the most significant module (subnetwork 1) included 46 nodes (all genes down-regulated) and 326

edges (Score=14.489) (*Figure 4C*). Centrality analysis of genes in the subnetwork1 showed that top 3 genes among genes obtained by each Centrality method were *FPR2* ($\log FC=-3.511$, $P=1.08e-74$), *GNG11* ($\log FC=-2.994$, $P=4.22e-110$) and *ADCY4* ($\log FC=-2.472$, $P=1.15e-99$) (*Table 3*), and the 3 genes were identified as essential genes in subnetwork1. The expression analysis based on GEPIA database showed that the 3 genes were significantly down-regulated expressed in LUSC tissues (*Figure 4D*), and which were consistent with the results obtained in this study. The expression analysis between major pathological stagings showed that *FPR2* ($F=2.76$ and $P=0.0419$) had significant difference between four stages, with higher expression in stage IV (*Figure 4E*). Survival analysis revealed that low mRNA expression of *FPR2* and *GNG11* resulted in a higher OS rate than high mRNA expression ($P=0.036$ and 0.049 , respectively) (*Figure 4F*). Pearson correlation analysis showed that mRNA expression between *FPR2* and *GNG11* ($R=0.44$ and $P=0$) had more similar pattern in LUSC and normal lung tissues (*Figure 4G*), which demonstrated that *FPR2* and *GNG11* possessed the stronger co-expression in LUSC and normal lung tissues. The findings indicate that *FPR2* and *GNG11* may sever as potential prognosticators of OS in LUSC patients.

Construction of ceRNA network

The relationships among DElncRNA, DEmiRNA and



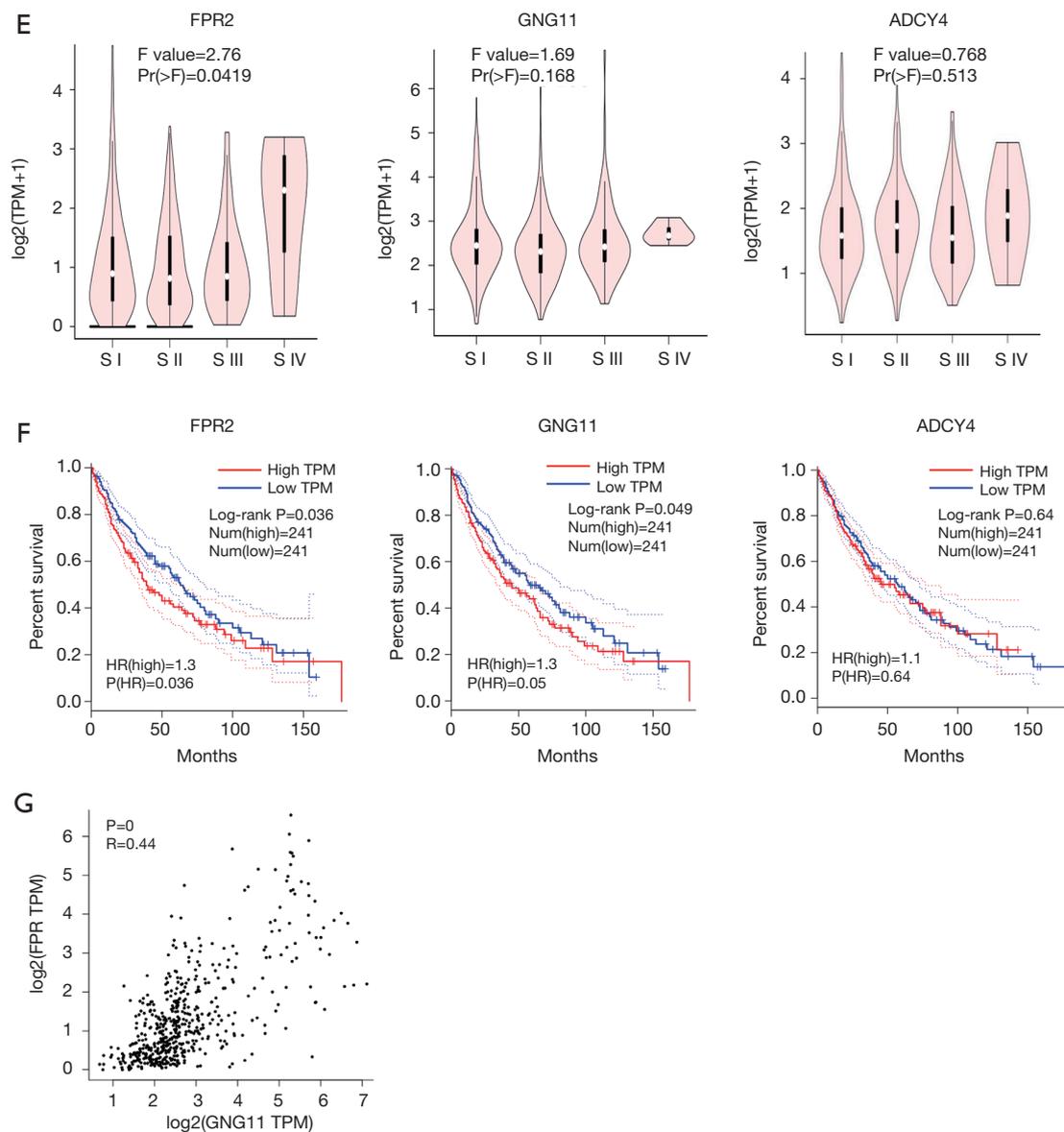


Figure 4 Molecular network and key gene analysis in LUSC. (A) Ten pathways and 97 genes were used to construct pathway-gene network. Red cycle represented gene, and blue triangle represented pathway. (B) PPI network. 239 genes including 232 down-regulated and 7 up-regulated genes were used to construct the PPI network. Nodes with red and black labels represented up-regulated and down-regulated genes, respectively. Red, green and blue nodes represented the genes in highly correlated modules by MCODE methods. (C) highly correlated module analysis showed that three subnetworks were identified, and the most significant subnetwork included 46 nodes and 326 edges. Each node represented one gene, and bigger nodes represented genes with more links. Each edge represented the interactive relationship between two genes. Thicker edges represented higher co-expression score among genes and deeper color edges (green to yellow to black) represented higher combined score among genes. (D) The expression of *FPR2*, *GNG11* and *ADCY4* mRNAs were significantly down-regulated in LUSC tissues. (E) The expression of *FPR2* had significant difference between major pathological stagings, with higher expression in stage IV. (F) Low mRNA expression of *FPR2* and *GNG11* was significantly associated with a higher overall survival rate of LUSC patients. (G) The mRNA expression between *FPR2* and *GNG11* had more similar pattern in LUSC and normal tissues. LUSC, lung squamous cell carcinoma; PPI, protein and protein interaction; MCODE, molecular complexity detection; TPM, transcripts per million; HR, hazard ratio.

Table 3 Top 3 genes with higher scores identified by centrality analyses

Rank	Gene	Subgraph	Degree	Eigenvector	Information	Betweenness	Closeness	Network
1	<i>FPR2</i>	810,463.75	34	0.2978219	10.2128935	822.15875	822.15875	32.771732
2	<i>GNG11</i>	771,008.3	31	0.29110122	9.952475	390.3095	390.3095	30.515686
3	<i>ADCY4</i>	594,200.06	25	0.255286	9.313109	257.04285	257.04285	23.600285

DEmRNAs were required before lncRNA-miRNAs-mRNAs ceRNA network was constructed. Using online miRcode tool, the relationships among 1,694 DElncRNAs and 113 DEmiRNAs in the turquoise module were evaluated, and finally 121 LUSC-specific DElncRNAs (92 up- and 29 down-regulated, *Table 4*) and 18 LUSC-specific DEmiRNAs (12 up- and 6 down-regulated, *Table 4*) were putatively identified to interact. Using online tools including TargetScan, miRDB and miRTarBase, we predicted the relationships among DEmiRNAs and their target DEmRNAs. Finally, 3 LUSC-specific target DEmRNAs (1 up- and 2 down-regulated, *Table 4*) were putatively identified to interact with 18 LUSC-specific DEmiRNAs.

On the basis of the above data, lncRNA-miRNA-mRNA ceRNA network was constructed and visualized using Cytoscape software. Based on ceRNA hypothesis and the expression levels of DElncRNA, DEmiRNAs and DEmRNAs, two lncRNA-miRNA-mRNA ceRNA networks including over-expressed (*Figure 5A*) and under-expressed (*Figure 5B*) networks were constructed. In over-expressed ceRNA network, 65 up-regulated DElncRNAs, 6 down-regulated DEmiRNAs and 1 up-regulated DEmRNAs were included and 111 edges were contained. In under-expressed ceRNA network, 23 down-regulated DElncRNAs, 12 up-regulated DEmiRNAs and 1 down-regulated DEmRNAs were included and 78 edges were contained.

OS analysis of LUSC-specific lncRNA signature

The relationships of OS in LUSC patients and DElncRNAs in ceRNA network were determined using KM estimate and LR test. According to $P < 0.05$ cut-off threshold, 8 DElncRNAs including *AL391152.1* [$P = 3.834 \times 10^{-3}$, HR(high) = 0.66346, 95% CI: 0.4967–0.8763], *CACNA2D3-AS1* [$P = 8.581 \times 10^{-3}$, HR(high) = 0.68679, 95% CI: 0.5149–0.9095], *KCNQ5-IT1* [$P = 0.013$, HR(high) = 0.70386, 95% CI: 0.5266–0.9287], *LNX1-AS1* [$P = 0.03775$, HR(high) = 1.34097, 95% CI: 1.016–1.79], *MAGI2-AS3* [$P = 0.01855$, HR(high) = 1.39845, 95% CI: 1.0569–1.8601], *POU6F2-AS2*

[$P = 9.601 \times 10^{-3}$, HR(high) = 0.69481, 95% CI: 0.5189–0.9147], *SOX2-OT* [$P = 0.03924$, HR(high) = 0.74685, 95% CI: 0.5612–0.9865] and *TTY16* [$P = 0.02715$, HR(high) = 1.3863, 95% CI: 1.0483–1.8538] were found to be related to OS of LUSC patients (*Figure S1*).

Establishment of lncRNA signature prognostic model

We used univariate regression analysis to identify DElncRNAs associated with the OS of LUSC patients. With $P < 0.05$ cut-off threshold, a group of 10 DElncRNAs including *TTY16* [$P = 0.00583$, HR(high) = 1.155836, 95% CI: 1.04276–1.28118], *ADAMTS9-AS2* [$P = 0.01317$, HR(high) = 1.15863, 95% CI: 1.03132–1.30165], *AC006238.1* [$P = 0.014560$, HR(high) = 1.16733, 95% CI: 1.03107–1.32160], *AC123595.1* [$P = 0.01636$, HR(high) = 1.16110, 95% CI: 1.02780–1.31168], *CACNA2D3-AS1* [$P = 0.01905$, HR(high) = 0.87871, 95% CI: 0.78878–0.97903], *KCNQ5-IT1* [$P = 0.02881$, HR(high) = 0.89105, 95% CI: 0.80350–0.98814], *POU6F2-AS2* [$P = 0.03166$, HR(high) = 0.92865, 95% CI: 0.86801–0.99352], *AL391152.1* [$P = 0.03411$, HR(high) = 0.90647, 95% CI: 0.82776–0.99267], *CHODL-AS1* [$P = 0.03806$, HR(high) = 0.87980, 95% CI: 0.77952–0.99297] and *LINC00462* [$P = 0.04096$, HR(high) = 0.93351, 95% CI: 0.87391–0.99718] was detected to have significantly prognostic value (<http://fp.amegroups.cn/cms/23ad548482adf5ccf686ce250556033f/tlcr.2019.09.13-4.pdf>). Furthermore, we found that 5 DElncRNAs (*TTY16*, *CACNA2D3-AS1*, *KCNQ5-IT1*, *POU6F2-AS2* and *AL391152.1*) were simultaneously identified to be related to OS in KM (LR test) and univariate regression analysis. The 5 DElncRNAs were fitted into the multivariate Cox regression model with Stepwise method. The result showed that a group of 3 DElncRNAs including *TTY16*, *POU6F2-AS2* and *CACNA2D3-AS1* had significantly prognostic value with OS of LUSC patients, and 3-lncRNA prognostic model was established ($P = 0.002891$). High lncRNA expression of *POU6F2-AS2* and *CACNA2D3-AS1*, and low mRNA expression of *TTY16* resulted in a higher OS in

Table 4 DERNA's in ceRNA network of LUSC

RNA	logFC	P value	FDR
LncRNA			
LINC00392	9.440110812	1.54E-14	9.88E-14
LINC00355	7.791573469	5.43E-23	6.90E-22
POU6F2-AS2	7.639818861	3.74E-54	2.29E-52
AL513123.1	7.295631459	9.06E-39	2.90E-37
HOTAIR	6.277782238	1.22E-17	1.04E-16
MIR137HG	6.227772241	4.95E-13	2.77E-12
MIR205HG	6.149099342	1.46E-58	1.13E-56
LINC00536	5.751355879	3.13E-14	1.96E-13
TTY20	5.58577463	2.12E-04	4.15E-04
TDRG1	5.5076554	1.34E-09	5.30E-09
SOX2-OT	5.475037306	7.74E-21	8.49E-20
LINC00462	5.374797677	2.26E-15	1.56E-14
AC006305.1	5.118874308	2.10E-34	5.34E-33
LSAMP-AS1	4.998808	9.45E-25	1.35E-23
LINC00461	4.897041384	8.40E-14	5.04E-13
PEX5L-AS1	4.885905232	4.68E-04	8.76E-04
LINC00518	4.818106321	6.29E-11	2.89E-10
UCA1	4.789191312	1.54E-15	1.08E-14
IL20RB-AS1	4.727889802	2.19E-16	1.66E-15
PART1	4.479053831	6.10E-17	4.87E-16
LINC00470	4.443384297	4.63E-16	3.38E-15
AL163952.1	4.38576249	4.12E-14	2.55E-13
LINC00051	4.333975408	1.10E-08	3.89E-08
KCNQ5-IT1	4.246609506	1.27E-17	1.08E-16
LINC00404	4.232371376	6.00E-05	1.28E-04
SOX21-AS1	4.146402175	2.14E-28	3.84E-27
LINC00308	4.053357554	2.86E-07	8.48E-07
LINC00488	3.978445763	2.48E-05	5.56E-05
AP002478.1	3.936160123	2.90E-15	1.97E-14
AC080129.1	3.926928079	4.21E-08	1.38E-07
GRM5-AS1	3.73674902	1.10E-10	4.88E-10
GDNF-AS1	3.643490224	4.20E-13	2.37E-12
LINC00320	3.586351022	7.76E-05	1.62E-04
AL391152.1	3.583026125	3.77E-31	7.96E-30
LNK1-AS1	3.545403022	5.17E-09	1.91E-08

Table 4 (continued)**Table 4** (continued)

RNA	logFC	P value	FDR
MLIP-IT1	3.483310114	6.08E-17	4.86E-16
LINC00391	3.464052328	2.83E-11	1.35E-10
C10orf91	3.413159709	3.59E-24	4.94E-23
AC109462.1	3.374590902	5.81E-12	2.96E-11
KIRREL3-AS3	3.331565793	1.76E-05	4.05E-05
CHODL-AS1	3.276032463	3.33E-10	1.41E-09
ADIPOQ-AS1	3.160141054	2.96E-03	4.92E-03
AC016912.1	3.08743503	7.95E-04	1.44E-03
LINC00520	3.079503876	7.79E-11	3.51E-10
C12orf77	3.022273593	3.30E-05	7.30E-05
LINC00052	2.997128874	2.65E-05	5.94E-05
AC100791.1	2.994926916	7.32E-08	2.33E-07
AL139385.1	2.99010878	3.68E-28	6.49E-27
DSCAM-AS1	2.985792179	3.52E-03	5.80E-03
AC025211.1	2.947705948	2.26E-18	2.06E-17
LINC00460	2.945188266	5.49E-08	1.78E-07
AC012640.1	2.940659458	2.40E-22	2.94E-21
LINC00458	2.87315071	6.69E-04	1.22E-03
LINC00492	2.867750469	1.36E-11	6.65E-11
AC092811.1	2.835233349	8.16E-09	2.93E-08
STEAP2-AS1	2.780879429	1.79E-15	1.25E-14
C11orf44	2.766012836	4.66E-06	1.17E-05
C2orf48	2.721907434	3.70E-22	4.46E-21
LINC00504	2.709918399	5.66E-10	2.33E-09
CACNA2D3-AS1	2.685950038	3.38E-08	1.12E-07
LINC00319	2.636763346	5.56E-08	1.80E-07
LINC00381	2.635007306	2.17E-08	7.39E-08
LINC00184	2.623125522	1.15E-12	6.28E-12
AC006372.1	2.616079799	1.35E-04	2.72E-04
AP000525.1	2.59642911	2.21E-16	1.67E-15
AC018781.1	2.586720267	1.27E-12	6.82E-12
TBL1XR1-AS1	2.580626632	3.64E-14	2.26E-13
AC022762.1	2.572878977	4.37E-09	1.63E-08
SMCR2	2.521422296	1.07E-14	6.93E-14
AC009061.1	2.518928136	2.67E-09	1.02E-08
CLDN10-AS1	2.509955326	1.44E-03	2.50E-03

Table 4 (continued)

Table 4 (continued)

RNA	logFC	P value	FDR
C17orf77	2.462406767	1.78E-03	3.06E-03
AC078802.1	2.453881668	1.21E-17	1.03E-16
AL136131.2	2.366865903	1.19E-05	2.82E-05
WNT5A-AS1	2.353778973	7.62E-14	4.59E-13
PVT1	2.349482579	2.70E-38	8.25E-37
LINC00485	2.256897083	1.07E-05	2.55E-05
AC078778.1	2.256314434	1.15E-28	2.11E-27
IGF2-AS	2.250318994	2.13E-05	4.83E-05
RRM1-AS1	2.241727965	4.80E-11	2.24E-10
AC079070.1	2.234359786	2.23E-05	5.05E-05
LINC00534	2.157428061	3.14E-05	6.97E-05
AC128709.1	2.131149708	9.37E-07	2.59E-06
AC092422.1	2.126433862	8.71E-06	2.11E-05
LINC00173	2.124251844	4.73E-14	2.90E-13
STEAP3-AS1	2.11620078	5.21E-12	2.67E-11
AC097717.1	2.10623016	1.46E-13	8.59E-13
MTUS2-AS1	2.099780402	2.59E-07	7.75E-07
AC002511.1	2.065525454	5.35E-05	1.15E-04
AC073352.1	2.049896234	1.28E-15	9.10E-15
LINC00525	2.046446311	2.80E-08	9.44E-08
C8orf49	2.046391517	1.02E-03	1.82E-03
KCNQ1-AS1	-2.034335916	3.00E-18	2.68E-17
AC006238.1	-2.133822774	2.43E-37	7.07E-36
LINC00469	-2.175169274	7.52E-15	4.93E-14
AC123595.1	-2.179056388	1.41E-34	3.62E-33
LINC00113	-2.190134588	8.01E-09	2.89E-08
TTY16	-2.194922826	5.23E-09	1.93E-08
LINC00472	-2.232889162	3.70E-39	1.20E-37
MAGI2-AS3	-2.290846084	6.84E-56	4.71E-54
TLR8-AS1	-2.345990308	3.50E-14	2.18E-13
AL139819.1	-2.376701055	1.81E-20	1.94E-19
BCRP3	-2.410412265	2.43E-51	1.35E-49
KIAA0087	-2.416352287	2.59E-16	1.95E-15
AC093510.2	-2.423740876	8.69E-42	3.11E-40
MED4-AS1	-2.437359227	2.26E-72	2.66E-70
NKX2-1-AS1	-2.511076899	3.48E-18	3.10E-17

Table 4 (continued)

Table 4 (continued)

RNA	logFC	P value	FDR
LINC00211	-2.672798113	4.11E-27	6.76E-26
AC025431.1	-2.68572414	3.16E-20	3.30E-19
ATP13A4-AS1	-2.744528113	1.40E-14	9.01E-14
C5orf64	-2.871131256	4.04E-57	2.94E-55
MYO16-AS1	-3.043363778	1.57E-31	3.41E-30
AC135178.1	-3.080541278	2.75E-45	1.15E-43
AP000438.1	-3.256244279	1.93E-57	1.43E-55
ADAMTS9-AS2	-3.624852381	2.63E-149	2.32E-146
LINC00163	-3.64214957	2.30E-45	9.78E-44
RMST	-3.882619917	6.84E-41	2.36E-39
ADAMTS9-AS1	-4.088690083	2.87E-121	1.33E-118
AC105206.1	-4.356484234	1.54E-91	3.31E-89
AL355388.1	-4.517353966	6.41E-105	1.82E-102
SFTA1P	-4.619645817	6.58E-113	2.42E-110
miRNA			
hsa-mir-31	5.603226184	2.45E-25	2.61E-24
hsa-mir-205	5.575113876	2.63E-55	9.01E-54
hsa-mir-301b	4.29013536	2.10E-28	2.61E-27
hsa-mir-210	4.244884194	9.62E-67	4.94E-65
hsa-mir-9	4.155804934	8.69E-19	6.37E-18
hsa-mir-383	3.246759952	7.18E-05	1.48E-04
hsa-mir-183	3.218906108	3.32E-31	4.55E-30
hsa-mir-96	2.681966878	7.56E-33	1.07E-31
hsa-mir-137	2.672308656	5.19E-06	1.27E-05
hsa-mir-25	2.204516321	1.24E-10	4.95E-10
hsa-mir-503	2.154171477	2.59E-20	2.03E-19
hsa-mir-182	2.000460902	2.85E-22	2.54E-21
hsa-mir-140	-2.055995223	5.29E-61	2.29E-59
hsa-mir-218	-2.130060686	1.14E-49	2.83E-48
hsa-mir-206	-2.166983977	2.62E-06	6.71E-06
hsa-mir-101	-2.380509751	1.16E-117	2.38E-115
hsa-mir-338	-2.631345685	6.13E-67	3.35E-65
hsa-mir-144	-3.09914232	9.64E-63	4.55E-61
mRNA			
PAX3	6.151717371	1.87E-17	7.95E-17
LIMCH1	-3.424019142	1.20E-117	1.47E-115
UTRN	-2.012223181	4.68E-83	2.25E-81

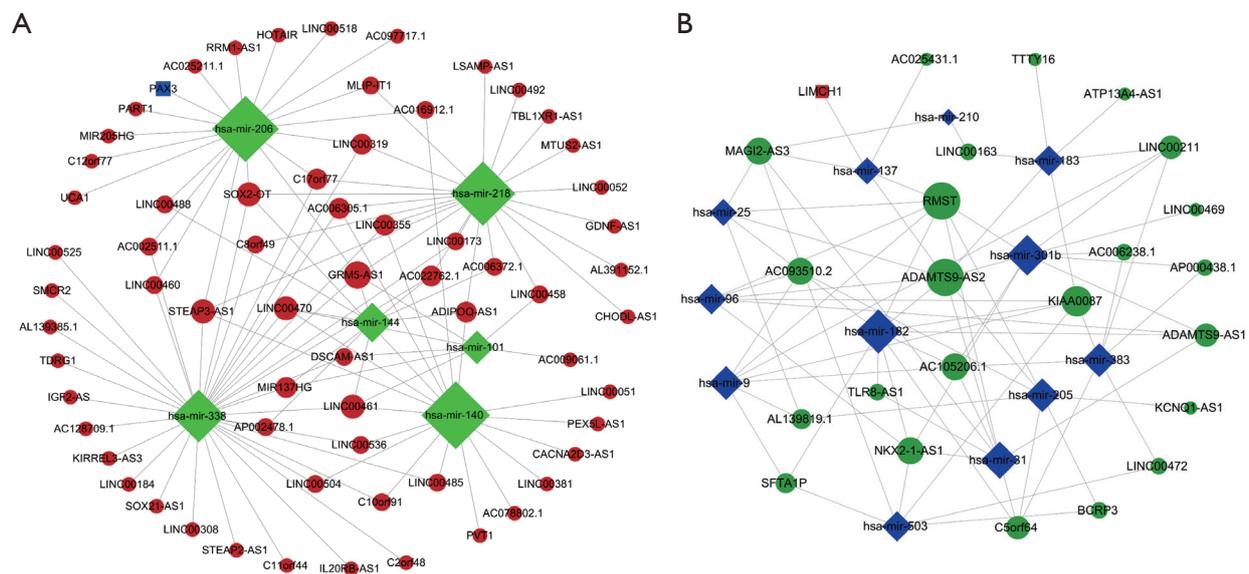


Figure 5 LncRNA-miRNA-mRNA ceRNA network. (A) Over-expressed network. Red circles represented up-regulated lncRNAs, green diamonds represented down-regulated miRNAs, and blue rectangles represented up-regulated mRNAs. (B) Under-expressed network. Green circles represented down-regulated lncRNAs, blue diamonds represented up-regulated miRNAs and red rectangles represented down-regulated mRNAs. ceRNA, competitive endogenous RNA.

patients (Figure 6A). With the risk score of each patient (<http://fp.amegroups.cn/cms/d5fa0921ac865b7ce793c7183465233d/tlcr.2019.09.13-5.pdf>), the LUSC patients were divided into high- and low-risk groups (Figure 6B,C). Comparing to low risk group, the mortality rate of patients in high-risk group was significantly higher [P=1e-05, HR(high) =1.8574, 95% CI: 1.4149–2.5091], and high-risk group was related with worse prognosis (Figure 6D). The ROC of 3-year survival correlation of the 3-lncRNA signature was analyzed, and the area under the curve (AUC) was computed. The AUC of 3-lncRNA signature was 0.629 (Figure 6E), which indicates its effectiveness as a prognostic biomarker in predicting OS of LUSC patients.

The expression levels of the 3 lncRNAs were analyzed between LUSC tissues and normal lung tissues and between high- and low- risk patient groups. *CACNA2D3-AS1* and *POU6F2-AS2* lncRNAs were significantly highly expressed in LUSC patients (P=2.136e-12 and <2.2e-16, separately, Figure 7A) with low risk scores (P<2.2e-16 and =1.959e-06, separately, Figure 7B). Whereas, *TTY16* lncRNA was significantly lowly expressed in LUSC patients (P=4.592e-07, Figure 7A) with high risk scores (P=1.429e-14, Figure 7B). Further, the expressions of three DELncRNAs between dead and alive groups were

observed, and showed that the expression patterns of three DELncRNAs between dead and alive groups were separately consistent with that between LUSC and normal tissues and between low and high-risk groups (Figure 7C). The expression of *CACNA2D-AS1* in alive group was significantly higher than that in dead group (P=0.002532, Figure 7C). The expressions of the 3 DELncRNAs in high- and risk-groups and survival status groups were showed in Figure 7D.

Discussion

LUSC is a common malignant cancer, and per year causes hundreds of thousands of deaths. Identifying specific diagnostic and prognostic biomarkers may contribute to improving survival rate among LUSC patients. To improve clinical outcomes, it is essential to identify LUSC-related prognostic signatures that predict those outcomes. Increasing evidence indicates that lncRNAs play key roles in many biological processes including cancer initiation and progression by ceRNA mechanism. In this study, we systematically analyzed the LUSC-related RNA-seq data from TCGA database by bioinformatics methods including WGCNA, GO and KEGG pathway analyses, PPI network

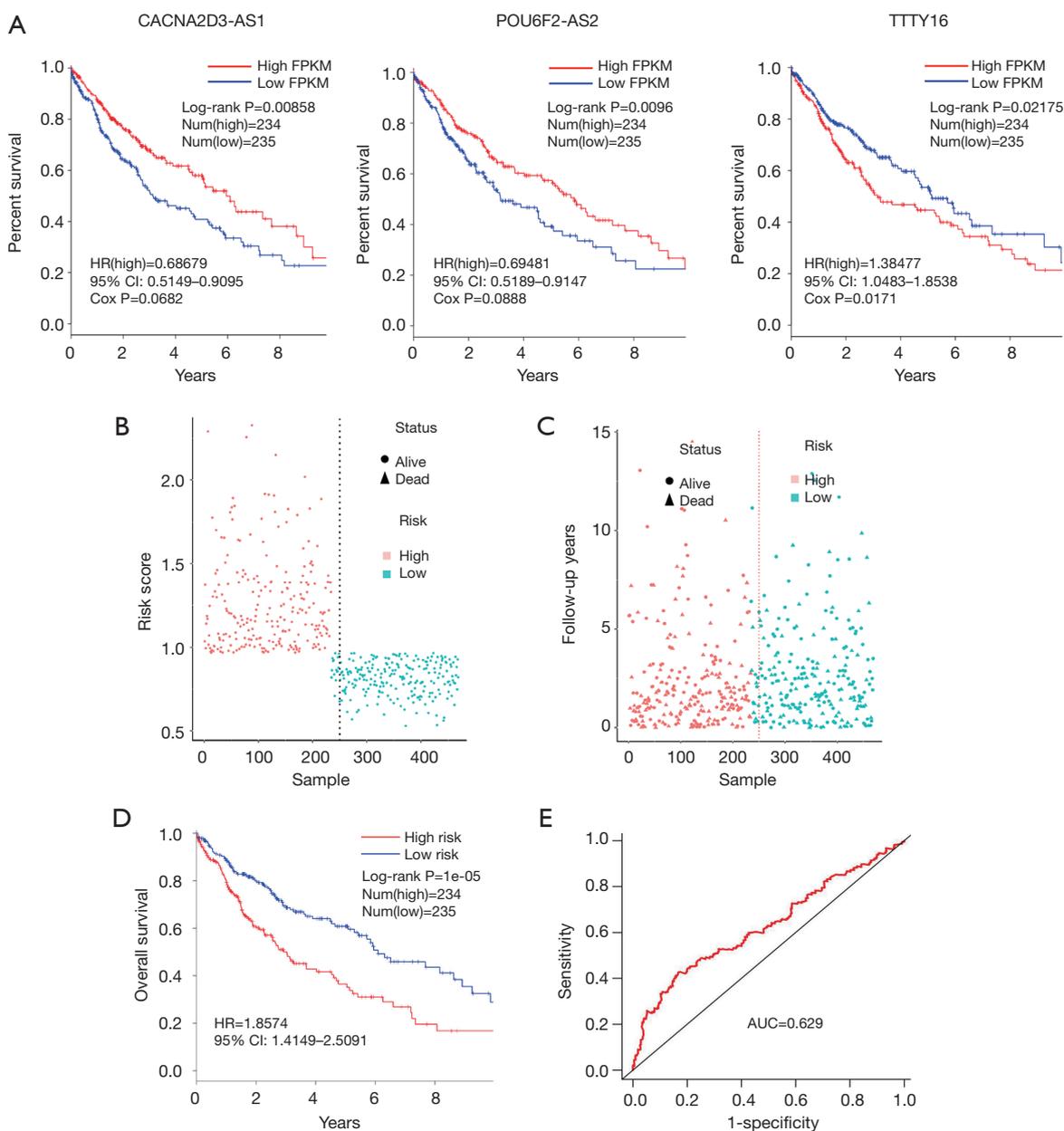


Figure 6 Kaplan-Meier survival curves and the outcome of 3-lncRNA signature. (A) Kaplan-Meier survival curves for 3 lncRNAs associated with overall survival of lung squamous cell carcinoma (LUSC) patients. Log-rank test was used to evaluate the survival differences between low- and high-expression groups. X-axis represented overall survival time (years) and Y-axis represented survival function. (B) The LUSC patients were divided into high- and low-risk groups according to risk scores of LUSC patients. (C) The survival information between high- and low-risk groups was plotted using ggplot2 package. (D) Survival analysis showed high risk group had worse prognosis in LUSC patients ($P=1e-05$). (E) ROC curve showed that AUC of 3-lncRNA prognostic model was 0.629. LUSC, lung squamous cell carcinoma; ROC, receiver operating characteristic; AUC, area under the curve; HR, hazard ratio; CI, confidence interval.

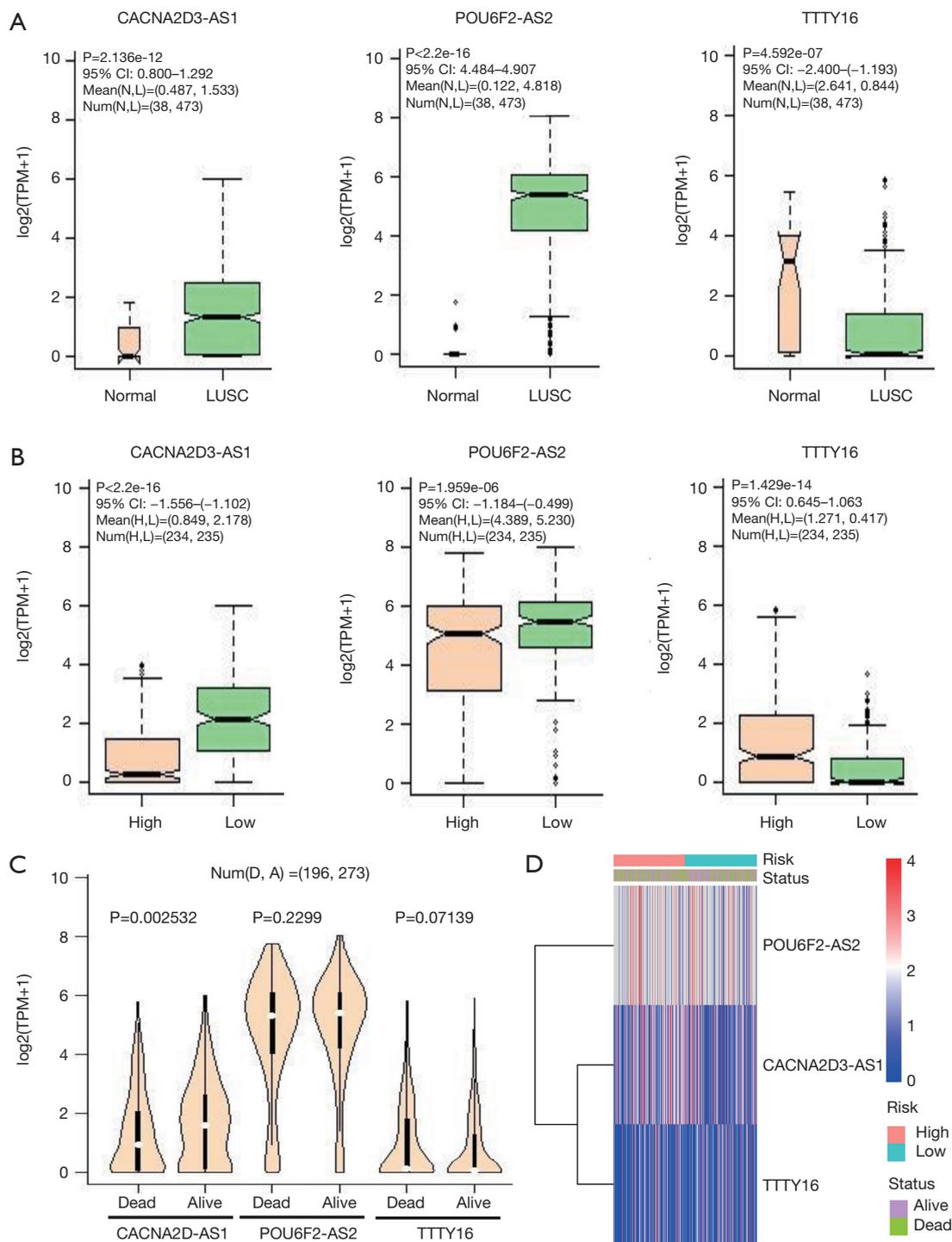


Figure 7 Expression pattern of the 3-lncRNA. (A) The expression pattern of the 3-lncRNA was in LUSC and normal tissues samples; (B) the expression pattern of 3-lncRNA was in high- and low-risk groups; (C) the expression distribution of 3-lncRNA was in alive and dead patients; (D) the expression level of 3-lncRNA was in experimental tissues samples. LUSC, lung squamous cell carcinoma; CI, confidence interval.

construction, ceRNA network construction and survival analysis, and identified the key genes and key lncRNA signature associated with OS in LUSC patients. Finally, two genes including *FPR2* and *GNG11* in PPI network, and three lncRNA signature including *CACNA2D3-AS1*, *POU6F2-AS2* and *TTTY16* in ceRNA network were identified to associate with OS in LUSC patients.

FPR2 is an important member of 7 transmembrane G-protein-coupled FPR family, and which was originally identified to play critical roles in host defense (40). Recently, *FPR2* has been reported to play a key role in human diseases such as cancers (41,42). For example, a study found that *FPR2* promoted invasion and metastasis of gastric cancer cells, and negative *FPR2* expression was associated with a higher OS of patients (41), which was consistent with our results. *GNG11* is a member of guanine nucleotide-binding protein (G protein) gamma family, and play a role in transmembrane signaling system. Recent studies showed that *GNG11* played roles in regulating cell growth and cellular senescence (43,44). So far, the functions of *GNG11* is entirely unknown in cancer. No studies have so far reported any association of two genes including *FPR2* and *GNG11* with LUSC. This study is the first to show aberrant expressions of *FPR2* and *GNG11* were identified to associate with OS of LUSC patients, and the low mRNA expressions of *FPR2* and *GNG11* resulted in a higher OS rate in LUSC patients.

In recent years, some studies have investigated ceRNA in LUSC, and constructed ceRNA network to identify some lncRNAs associated with OS of LUSC patients (9,26,27). For example, Sui et al. constructed lncRNA regulatory ceRNA by integrating analyzing RNA-seq data from TCGA database using DEGA and ceRNA network, and constructed a 22 lncRNAs multivariate Cox regression model and finally identified two lncRNAs including *FMO6P* and *PRR26* had significantly prognostic value (26). In addition, Ning et al. also used LUSC-related from TCGA database to construct lncRNA-miRNA-mRNA ceRNA network, and identified six ceRNAs (*PLAU*, *miR-31-5p*, *miR-455-3p*, *FAM83A-AS1*, *MIR31HG*, and *MIR99AHG*) significantly correlated with survival by survival analysis based on KM method (9). These LUSC-related ceRNA studies provide novel knowledge for a better understanding the ceRNA interactive mechanism in LUSC biology, and contribute to improving the diagnosis and prognosis of LUSC patients. However, few lncRNA-related ceRNA network studies considered the co-expression among ceRNAs. In this study, we firstly used WGCNA to identify the co-expression

modules in LUSC, and further constructed lncRNA-miRNA-mRNA ceRNA network based on co-expression genes in the most significant turquoise module according to ceRNA theory. Furthermore, KM (LR test) and univariate Cox regression analysis revealed that 5 DELncRNAs in ceRNA network (*TTTY16*, *CACNA2D3-AS1*, *KCNQ5-IT1*, *POU6F2-AS2* and *AL391152.1*) were simultaneously identified to associate with OS of LUSC patients. Multivariate Cox regression analysis showed significant prognostic value of 3 of those lncRNAs (*CACNA2D3-AS1*, *POU6F2-AS2* and *TTTY16*) in the OS of LUSC patients. A cumulative risk score of the 3-lncRNA showed that 3-lncRNA signature could independently predict OS in LUSC patients.

Among identified 3-lncRNA signature, *POU6F2-AS2* lncRNA is an antisense lncRNA, and has so far been reported to play role in esophageal squamous cell carcinoma by modulating DNA repair (45). However, no studies reported the association between *POU6F2-AS2* and OS of cancer-related patients. In our study, we found that *POU6F2-AS2* was highly expressed in LUSC tissue, and LUSC patients with high expression possessed low risk scores. Similar to *POU6F2-AS2*, *CACNA2D3-AS1* is an antisense lncRNA, and no studies so far has reported any association of *CACNA2D3-AS1* with cancer. In our study, we noticed that *CACNA2D3-AS1* in ceRNA network was up-regulated and competed with the down-regulated hsa-mir-140. Some previous studies showed that hsa-mir-140 often functioned as a tumor suppressor in cancer (46-48). Down-regulated hsa-mir-140 has been suggested to enhance cell proliferation and invasion in colorectal carcinoma and squamous cell LC (46,47), and low expression of hsa-mir-140 is associated with poor prognosis of spinal chordoma (49). This study firstly showed aberrant expression of *CACNA2D3-AS1* in LUSC, and indicated that *CACNA2D3-AS1* might sever as a potentially prognostic marker in LUSC. In addition, we observed that *TTTY16* in ceRNA network was down-regulated in LUSC, and *TTTY16* with low expression competed with up-regulated hsa-mir-183. *TTTY16* has so far been reported to be associated with cancer (5), and a 14-gene expression signature including *TTTY16* was related to the prognoses of LUSC patients (5). Previous studies also showed that hsa-mir-183 was highly expressed in cancer, and promoted tumor cell growth and migration (50,51), which indicated that the high expression of hsa-mir-183 was related to the poor prognosis. So far, no studies about *TTTY16* regulating cancer initiation and progression by competing with has-

mir-183 was reported. In this study, we are the first to report the role of *TTY16* by competing with hsa-mir-183 in LUSC, and low expression of *TTY16* was associated with good prognosis in LUSC. This study findings indicate that *CACNA2D3-AS1* and *POU6F2-AS2* may play roles as LUSC suppressors, and *TTY16* may function as LUSC promoter.

Despite the findings in clinical implications, some limitations must be noted. Firstly, the findings are obtained based on the TCGA database by pure bioinformatics methods. Although some findings have been validated by GEPIA database, the findings must be verified by experimental methods. Secondly, the 3 prognosticators of 3-lncRNA signature need be evaluated by a median follow-up duration (3 to 5 years).

Conclusions

Taken together, three lncRNAs were identified as a potentially prognostic lncRNA signature for LUSC patients. The findings provide novel insights into the lncRNA-related regulatory mechanism by ceRNA network in LUSC. Further, more functional studies are required to elucidate the lncRNA-related molecular mechanisms underlying LUSC.

Acknowledgments

Funding: This work was supported by National Natural Science Foundation of China (no 31660655), Yunnan Province Applied Basic Research Projects (nos 2016FB146, 2017FB042, 2018FE001) and Fund from Health and Family Planning Commission of Yunnan Province (no 2017NS261).

Footnote

Conflicts of Interest: The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

References

1. Bray F, Ferlay J, Soerjomataram I, et al. Global cancer

statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2018;68:394-424.

2. Blandin Knight S, Crosbie PA, Balata H, et al. Progress and prospects of early detection in lung cancer. *Open Biol* 2017;7. doi: 10.1098/rsob.170070.
3. Ling DJ, Chen ZS, Liao QD, et al. Differential effects of MTSS1 on invasion and proliferation in subtypes of non-small cell lung cancer cells. *Exp Ther Med* 2016;12:1225-31.
4. Gandara DR, Hammerman PS, Sos ML, et al. Squamous cell lung cancer: from tumor genomics to cancer therapeutics. *Clin Cancer Res* 2015;21:2236-43.
5. Li J, Wang J, Chen Y, et al. A prognostic 4-gene expression signature for squamous cell lung carcinoma. *J Cell Physiol* 2017;232:3702-13.
6. Tanoue LT, Detterbeck FC. New TNM classification for non-small-cell lung cancer. *Expert Rev Anticancer Ther* 2009;9:413-23.
7. Nesbitt JC, Putnam JB, Jr., Walsh GL, et al. Survival in early-stage non-small cell lung cancer. *Ann Thorac Surg* 1995;60:466-72.
8. Campbell JD, Alexandrov A, Kim J, et al. Distinct patterns of somatic genome alterations in lung adenocarcinomas and squamous cell carcinomas. *Nat Genet* 2016;48:607-16.
9. Ning P, Wu Z, Hu A, et al. Integrated genomic analyses of lung squamous cell carcinoma for identification of a possible competitive endogenous RNA network by means of TCGA datasets. *PeerJ* 2018;6:e4254.
10. Cancer Genome Atlas Research N. Comprehensive genomic characterization of squamous cell lung cancers. *Nature* 2012;489:519-25.
11. Cancer Genome Atlas Research N. Comprehensive molecular profiling of lung adenocarcinoma. *Nature* 2014;511:543-50.
12. Mercer TR, Dinger ME, Mattick JS. Long non-coding RNAs: insights into functions. *Nat Rev Genet* 2009;10:155-9.
13. Fang Y, Fullwood MJ. Roles, Functions, and Mechanisms of Long Non-coding RNAs in Cancer. *Genomics Proteomics Bioinformatics* 2016;14:42-54.
14. Arun G, Diermeier SD, Spector DL. Therapeutic Targeting of Long Non-Coding RNAs in Cancer. *Trends Mol Med* 2018;24:257-77.
15. Chen L, Dzakah EE, Shan G. Targetable long non-coding RNAs in cancer treatments. *Cancer Lett* 2018;418:119-24.
16. Wang L, Yu Z, Sun S, et al. Long non-coding RNAs: potential molecular biomarkers for gliomas diagnosis and

- prognosis. *Rev Neurosci* 2017;28:375-80.
17. Deng H, Wang JM, Li M, et al. Long non-coding RNAs: New biomarkers for prognosis and diagnosis of colon cancer. *Tumour Biol* 2017;39:1010428317706332.
 18. Salmena L, Poliseno L, Tay Y, et al. A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? *Cell* 2011;146:353-8.
 19. Conte F, Fisco G, Chiara M, et al. Role of the long non-coding RNA PVT1 in the dysregulation of the ceRNA-ceRNA network in human breast cancer. *PLoS One* 2017;12:e0171661.
 20. Fan CN, Ma L, Liu N. Systematic analysis of lncRNA-miRNA-mRNA competing endogenous RNA network identifies four-lncRNA signature as a prognostic biomarker for breast cancer. *J Transl Med* 2018;16:264.
 21. Liu H, Zhang Z, Wu N, et al. Integrative Analysis of Dysregulated lncRNA-Associated ceRNA Network Reveals Functional lncRNAs in Gastric Cancer. *Genes (Basel)* 2018;9. pii: E303.
 22. Lu M, Xu X, Xi B, et al. Molecular Network-Based Identification of Competing Endogenous RNAs in Thyroid Carcinoma. *Genes (Basel)* 2018;9. pii: E44.
 23. Cao Y, Wang P, Ning S, et al. Identification of prognostic biomarkers in glioblastoma using a long non-coding RNA-mediated, competitive endogenous RNA network. *Oncotarget* 2016;7:41737-47.
 24. Du Z, Sun T, Hacisuleyman E, et al. Integrative analyses reveal a long noncoding RNA-mediated sponge regulatory network in prostate cancer. *Nat Commun* 2016;7:10982.
 25. Zhou M, Wang X, Shi H, et al. Characterization of long non-coding RNA-associated ceRNA network to reveal potential prognostic lncRNA biomarkers in human ovarian cancer. *Oncotarget* 2016;7:12598-611.
 26. Sui J, Xu SY, Han J, et al. Integrated analysis of competing endogenous RNA network revealing lncRNAs as potential prognostic biomarkers in human lung squamous cell carcinoma. *Oncotarget* 2017;8:65997-6018.
 27. Tang RX, Chen WJ, He RQ, et al. Identification of a RNA-Seq based prognostic signature with five lncRNAs for lung squamous cell carcinoma. *Oncotarget* 2017;8:50761-73.
 28. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 2008;9:559.
 29. Ding M, Li F, Wang B, et al. A comprehensive analysis of WGCNA and serum metabolomics manifests the lung cancer-associated disordered glucose metabolism. *J Cell Biochem* 2019;120:10855-63.
 30. Di Y, Chen D, Yu W, et al. Bladder cancer stage-associated hub genes revealed by WGCNA co-expression network analysis. *Hereditas* 2019;156:7.
 31. Zhai X, Xue Q, Liu Q, et al. Colon cancer recurrence-associated genes revealed by WGCNA coexpression network analysis. *Mol Med Rep* 2017;16:6499-505.
 32. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 2010;26:139-40.
 33. Szklarczyk D, Franceschini A, Wyder S, et al. STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res* 2015;43:D447-52.
 34. Shannon P, Markiel A, Ozier O, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 2003;13:2498-504.
 35. Bader GD, Hogue CW. An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics* 2003;4:2.
 36. Yu G, Wang LG, Han Y, et al. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* 2012;16:284-7.
 37. Tang Y, Li M, Wang J, et al. CytoNCA: a cytoscape plugin for centrality analysis and evaluation of protein interaction networks. *Biosystems* 2015;127:67-72.
 38. Tang Z, Li C, Kang B, et al. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Res* 2017;45:W98-W102.
 39. Chandrashekar DS, Bashel B, Balasubramanya SAH, et al. UALCAN: A Portal for Facilitating Tumor Subgroup Gene Expression and Survival Analyses. *Neoplasia* 2017;19:649-58.
 40. Li L, Chen K, Xiang Y, et al. New development in studies of formyl-peptide receptors: critical roles in host defense. *J Leukoc Biol* 2016;99:425-35.
 41. Hou XL, Ji CD, Tang J, et al. FPR2 promotes invasion and metastasis of gastric cancer cells and predicts the prognosis of patients. *Sci Rep* 2017;7:3153.
 42. Xiang Y, Yao X, Chen K, et al. The G-protein coupled chemoattractant receptor FPR2 promotes malignant phenotype of human colon cancer cells. *Am J Cancer Res* 2016;6:2599-610.
 43. Hossain MN, Sakemura R, Fujii M, et al. G-protein gamma subunit GNG11 strongly regulates cellular senescence. *Biochem Biophys Res Commun* 2006;351:645-50.
 44. Takauji Y, Kudo I, En A, et al. GNG11 (G-protein

- subunit gamma 11) suppresses cell growth with induction of reactive oxygen species and abnormal nuclear morphology in human SUSM-1 cells. *Biochem Cell Biol* 2017;95:517-23.
45. Liu J, Sun X, Zhu H, et al. Long noncoding RNA POU6F2-AS2 is associated with oesophageal squamous cell carcinoma. *J Biochem* 2016;160:195-204.
46. Huang H, Wang Y, Li Q, et al. miR-140-3p functions as a tumor suppressor in squamous cell lung cancer by regulating BRD9. *Cancer Lett* 2019;446:81-9.
47. Zhao Z, Liu W, Li J. miR-140-5p inhibits cell proliferation and invasion in colorectal carcinoma by targeting SOX4. *Oncol Lett* 2019;17:2215-20.
48. Zhou Y, Wang B, Wang Y, et al. miR-140-3p inhibits breast cancer proliferation and migration by directly regulating the expression of tripartite motif 28. *Oncol Lett* 2019;17:3835-41.
49. Zou MX, Huang W, Wang XB, et al. Identification of miR-140-3p as a marker associated with poor prognosis in spinal chordoma. *Int J Clin Exp Pathol* 2014;7:4877-85.
50. Bi DP, Yin CH, Zhang XY, et al. MiR-183 functions as an oncogene by targeting ABCA1 in colon cancer. *Oncol Rep* 2016;35:2873-9.
51. Zhang L, Quan H, Wang S, et al. MiR-183 promotes growth of non-small cell lung cancer cells through FoxO1 inhibition. *Tumour Biol* 2015;36:8121-6.

Cite this article as: Hu J, Xu L, Shou T, Chen Q. Systematic analysis identifies three-lncRNA signature as a potentially prognostic biomarker for lung squamous cell carcinoma using bioinformatics strategy. *Transl Lung Cancer Res* 2019;8(5):614-635. doi: 10.21037/tlcr.2019.09.13

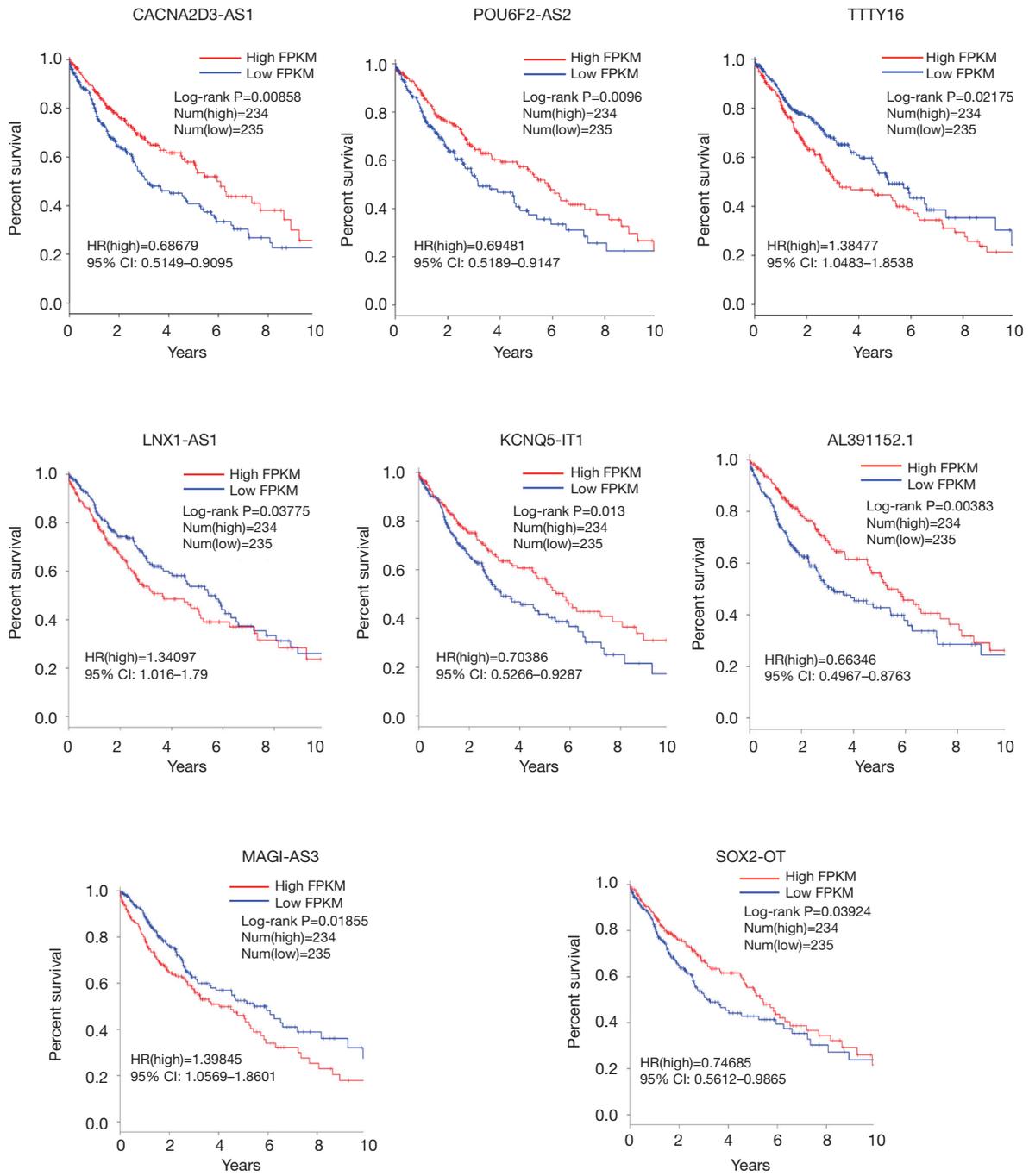


Figure S1 The relationships of lncRNAs and overall survival of lung squamous cell carcinoma (LUSC) patients.